·公众健康信息学·

基于 1D-ICNN 的高维度数据下老年自评健康预测方法

李 玥,张承蒙,黄成烨,索浩宇,胡新悦,刘 娜,张雅璐,陈 功
(北京大学人口研究所,北京 100871)

摘要:老年人自评健康是反映老年人身体健康状态的重要因子,对提高老年人健康水平提供参考具有重要意义。为了解影响我国农村老年人自评健康的主要因素并实现精准地预测,本研究基于 2022 年湖南省岳阳县养老需求调研数据,首先探究了不同影响因素对老年人自评健康的作用机制;然后基于显著影响因素,在面向高维度数据特征的情况下,提出一种基于交叉熵和变学习率的改进一维卷积神经网络(1D-ICNN)用于构建老年人自评健康预测模型,以解决 1D-CNN 容易出现预测不准确和不稳定等问题。本研究显示,老年人自评健康与文化程度、政治面貌、婚姻状况、职业、年收入等因素有关;在较高维度数据特征情况下,1D-ICNN 模型具有较好的预测效果。该方法的应用和普及能够为准确预测老年人健康状况、实现"健康老龄化"提供实证依据。

关键词:老年人;自评健康;一维卷积神经网络;预测模型
中图分类号:TP399 **文献标识码**:A **DOI**:10.3969/j.issn.1006-1959.2024.14.005
文章编号:1006-1959(2024)14-0025-08

Self-rated Health Prediction Method for the Elderly Based on 1D-ICNN High-dimensional Data

LI Yue,ZHANG Cheng-meng,HUANG Cheng-ye,SUO Hao-yu,HU Xin-yue,LIU Na,ZHANG Ya-lu,CHEN Gong
(Institute of Population Research,Peking University,Beijing 100871,China)

Abstract:The self-rated health of the elderly is an important factor to reflect the health status of the elderly, and it is of great significance to provide reference for improving the health level of the elderly. In order to understand the main factors affecting the self-rated health of the rural elderly in China and achieve accurate prediction, this study first explored the mechanism of different influencing factors on the self-rated health of the elderly based on the survey data of the elderly care demand in Yueyang County, Hunan Province in 2022. Then, based on the significant influencing factors, an improved one-dimensional convolutional neural network (1D-ICNN) based on cross entropy and variable learning rate is proposed to construct a self-rated health prediction model for the elderly in the case of high-dimensional data features, so as to solve the problems of inaccurate prediction and instability of 1D-CNN. This study shows that the self-rated health of the elderly is related to factors such as education level, political outlook, marital status, occupation and annual income. In the case of higher dimensional data features, the 1D-ICNN model has better prediction results. The application and popularization of this method can provide an empirical basis for accurately predicting the health status of the elderly and achieving "healthy aging".

Key words:Elderly;Self-rated health;One-dimensional convolutional neural network;Prediction model

人口老龄化已成为全世界人口发展的必然趋势,我国已成为老龄化速度最快的国家之一[1]。在我国城乡二元结构体制下,城乡之间的户籍壁垒、资源配置差异使得城镇居民和农村居民在认知和经历中都存在显著差异。根据第七次人口普查数据,截至 2020 年 11 月,岳阳县常住人口 561 888 人,60 岁以上、65 岁以上、80 岁以上老年人分别占 21.68%(全国占比 18.70%)、15.98%(全国占比 13.50%)和 2.97%。人口老龄化现象日趋严重,如何评价老年人健康是急需解决的问题[2]。同时,湖南省岳阳县作为我国农村创业创新典型县,其农村老年人的健康状况有其自身特色。在此背景下,对老年人健康问题进行深入研究有着十分重要的意义和价值。

评价老年人健康状况的指标较多,自评健康是调查者根据自身的身体、心理、社会功能等各方面综合情况对自身健康状况的主观评价与估计。自评健康在调查中经常运用和容易测量,目前已成为国际上运用广泛的健康状况测量方法之一[3]。国内外学者对关于老年人自评健康的问题开始受到学界的重视,主要集中对老年人自评健康的影响因素分析和预测方面[4]。

在探索影响老年人自评健康的决定因素方面,有研究[5]对影响老年人健康行为进行了全面的探析,发现老年人的健康生活方式与积极生活态度、健

基金项目:1. 中国工程院战略研究与咨询项目(编号:2022-XBZD-30);2.国家社会科学基金青年项目(编号:22CRK005)
作者简介:李玥(1994.8-),女,辽宁营口人,博士,助理研究员,主要从事智慧养老和福祉科技研究
通讯作者:陈功(1972.8-),男,北京人,博士,教授,主要从事社会老年学、养老服务、残疾和老龄健康等研究

康行为、对心理健康状态的关注、对疾病的预防，以及环境因素等密切相关。适量饮酒有益于身体健康[6]，因为饮酒可以降低某些心血管疾病的死亡概率。目前对影响老年人自评健康的因素主要集中在人口学特征、生活方式、患病情况、社会经济等[7]。人口学因素包括性别、年龄、受教育程度、婚姻状况等方面；生活方式因素，例如生活行为特征和饮食习惯都与老年人自评健康密切相关；患病情况与老年人自评健康也有显著的相关关系；社会因素包括生活环境、经济状况，社会参与等也会对老年人的健康状况产生影响[8]。影响老年人自评健康信息数据中存在变量冗余问题，这将降低预测有效性的同时造成模型的过拟合。很多学者使用单因素和多因素分析方法探索影响老年人自评健康的显著影响因素。单因素分析利用假设检验的方法来判断影响因素是否确实能解释因变量的变动，可以很容易地应对高维数据，结果具有良好的可解释性。目前使用较多的是卡方检验和方差分析。卡方检验用于研究分类变量与分类变量之间的差异关系，方差分析用于分析分类变量和定量变量之间差异关系[9]。

在老年人自评健康预测方面，机器学习在处理复杂数据问题时可以获得较好的精确度[10]。相比传统统计分析方法，基于分类算法的分析更加高效、客观，能够进一步支持健康的预测与预警。但随着影响老年人自评健康数据特征维度日益增加，传统分类算法的预测精度不高，很容易出现过拟合问题。因此，需要通过寻找更合适的分类算法提高老年人自评健康预测准确度。深度学习为解决高维度数据预测问题提供了新思路，其中 1D-CNN 是采用一维卷积核进行卷积操作，不仅能面向高维度数据省略掉复杂的人工特征提取工作，还能通过多层卷积操作提取到传统特征工程所无法提取到的抽象特征，但是 1D-CNN 模型的预测性能受到模型结构和参数设置等影响。随着网络深度和训练数据的增加，固定学习率难以适应网络的学习过程。在学习率优化方面，迭代过程中通常采用人工调整学习率、指数衰减、自适应参数等学习率变化方法。有研究[11]通过最大化局部似然估计来自动调整学习率，来防止学习率的波动；也有研究针对神经网络收敛性能较慢的问题[12]，在泰勒公式的基础上，提出自适应学习率的计算方法，结果表明该模型迭代次数明显少于基于固定学习率方法。

本研究拟探讨湖南省岳阳县农村老年人自评健康的影响因素并实现准确预测，对于提高老年人身心健康水平以及推进社会发展，具有重要的理论和实践意义。

1 数据来源

为了解我国农村老人的养老现状及需求，北京大学人口研究所于 2022 年在湖南省岳阳县开展农村养老服务专题调研实践活动。由于农村老年人居住分散、调研难度大，且调研时间有限，本研究选择通过分层抽样法进行研究以提高样本的代表性。删除有空缺的数据后，最终获得有效样本 369 份。根据研究需要，选取老年人自评健康作为因变量，即问卷中问题“您认为您的健康状况怎样？”选项为：很好、好、一般、不好、很不好，分别赋值为 1~5。选取与老年人自评健康相关的指标变量，包括：基本信息、家庭状况、生活方式、网络使用情况、养老需求等方面。老年人自评健康影响因素中的基本信息及赋值情况见表 1。

由于老年人自评健康选择为“不好”“很不好”“很好”“好”的样本数量较少，因此将“不好”“很不好”合并为“差”，将“很好”“好”合并为“好”。369 名农村老年人中，有 116 人（31.44%）表示自评健康为“好”，赋值为 1；121 人（32.79%）表示自评健康为“一般”，赋值为 2；132 人（35.77%）表示自评健康为“差”，赋值为 3。

从性别段来看，其中男 190 人，占比 51.49%，女 179 人，占比 48.51%；从年龄段来看，70~79 岁 202 人，占比为 54.74%。其他年龄段，60~69 岁 100 人，占比 27.10%，80~89 岁 56 人，占比 15.18%，90 岁及以上 11 人，占比 2.98%；从教育水平上看，小学学历 176 人，占比 47.70%，说明在本次调查中农村老年群体总体学历水平较低；从婚姻状况来看，与配偶居住的老年人 260 人，占比 70.46%，丧偶老年人 92人，占比 24.93%；从年收入情况来看，年收入 3000 元以下 126 人，占比 34.15%，低收入群体较多。

2 基于 1D-ICNN 的老年人自评健康预测模型

1D-ICNN 模型的输入向量是一维的，卷积层的卷积核和池化层的滤波器都相应变成一维，同时特征图也是一维的向量[13]，这样可以减少学习的参数数量，从而提高模型训练学习的效率。1D-ICNN 模型的基本结构见图 1。

表 1 老年人自评健康影响因素基本信息及赋值情况

特征变量	赋值说明	均值	标准差
性别	1=男;2=女	1.485	0.500
年龄	1=60~69 岁;2=70~79 岁;3=80~89 岁;4=90 岁及以上	1.940	0.735
户籍类型	1=农业;2=非农业;3=统一居民户	1.306	0.496
是否少数民族	1=是;2=否	1.003	0.052
文化程度	1=未上过学;2=小学;3=初中;4=高中/中专/职高;5=大学专科;6=本科及以上	2.379	0.942
政治面貌	1=群众;2=中共党员;3=无党派人士	1.209	0.426
婚姻状况	1=已婚,与配偶一同居住;2=已婚,没有与配偶一同居住;3=分居;4=离异;5=丧偶;6=从未结婚	2.095	1.771
工作情况	1=全职工作;2=兼职工作;3=不工作(退休)	1.702	0.481
职业	1=国家机关、党群组织、企业、事业单位;2=专业技术人员;3=办事人员和有关人员;4=商业、服务业人员;5=农、林、牧、渔、水利业生产人员;6=生产、运输设备操作人员;7=不便分类的工作	4.130	1.648
年收入	1=3000 元以下;2=3000~4999 元;3=5000~9999 元;4=10 000~19 999 元;5=20 000 元以上	2.938	1.678
年度人情支出	1=3000 元以下;2=3000~4999 元;3=5000~9999 元;4=10 000~19 999 元;5=20 000 元以上	2.553	1.388
宗教信仰	1=不信仰任何宗教;2=佛教;3=伊斯兰教;4=基督教;5=天主教;6=道教;7=其他宗教	1.182	0.822

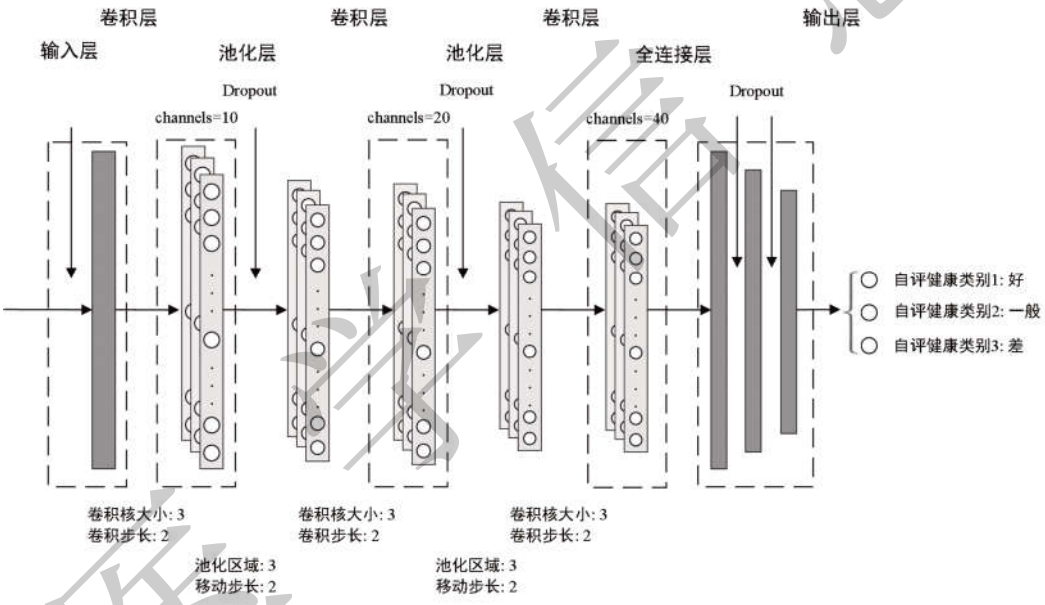

图 1 基于 1D-ICNN 模型的老年人健康自评状况预测模型

在输入层,首先基于卡方分析筛选的主要影响因素特征数据，对数据中的单选题进行 one-hot 编码后,共计生成 70 个特征维度数据用于建模;在卷积层,采用卷积核大小为 3,卷积步长为 2 的卷积操作,激活函数采用 ReLU 函数;在池化层,采用最大池化法，池化区域为 3，移动步长为 2 进行特征压缩。为了避免过拟合,提高模型的泛化能力,在最大池化后应用比例为 0.2 的随机失活操作(Dropout);全连接层通过对最后 Dropout 层输出的一维特征进行特征展开，提升模型的训练速度；最后通过 softmax 分类器预测老年人自评健康。网络模型中的结构设计和参数优化是预测问题的关键，本次 1D-ICNN 模型构建及优化策略如下：

卷积层：卷积层的作用是通过多个卷积核进行数据特征提取,具有权值共享和局部连接的优势。为了使模型学习更多的特征，通常采用多卷积核进行特征提取,有效降低模型的复杂程度。卷积运算可以表示为：

$$x_j^l = f\left(\sum_{i \in M_j} x_j^{l-1} * W_{ij}^l + b_j^l\right) \quad (1)$$

式中,* 表示卷积运算;l 表示当前网络的层数,第 l 层是特指卷积层;x_j^l 和 x_j^{l-1} 分别表示第 l 层和第

(l-1)层的第 j 个卷积核对应的特征向量；M_j^l 表示第 l 层第 j 个卷积核的视野域；W_{ij}^l 表示第 l 层的第 j 个卷积核的第 i 个加权值；b_j^l 是第 l 层的第 j 个卷积核对应的偏置；f(·)表示激活函数。

池化层：池化层具有二次特征提取的作用，主要在去除冗余特征的同时可以保留关键特征信息，有助于减少后续卷积操作的计算量。池化运算可以表示为：

$$r_j^p=f\left(\beta_j^p\times down\left(r_j^{p-1}\right)+b_j^p\right) \tag{2}$$

式中，down(·)表示池化函数；β_j^p 表示第 p 层的第 j 个特征图的加权值，第 p 层是特指池化层；b_j^p 是第 p 层的第 j 个特征图的偏置；r_j^p 为第 p 层池化层的第 j 个池化核对应的特征图。

全连接层和输出层：为保证模型可以最大程度地学习数据特征中的隐含知识，全连接层采用与上一层所有神经元进行连接的方式，同样也包含了线性操作和非线性操作，计算为：

$$z^q=f(z^{q-1}W^q+b^q) \tag{3}$$

式中，z^q 是第 q 层输出的特征图，即为全连接层输出的特征图；z^{q-1} 是第(q-1)层的输出特征图，即为上一层卷积和池化后输出的特征图；W^q 是第 q 层特征图 z^q 连接到 z^{q-1} 的权重；b^q 是第 q 层的偏置。如果研究任务为多分类问题，则输出层一般为 softmax 输出层，计算为：

$$p_k=\frac{\exp(z_k)}{\sum_{k=1}^{K}\exp(z_k)} \tag{4}$$

式中，p_k 表示当前输入数据属于第 k 类的概率；k 表示分类器的类别索引；K 是类别个数。z_k 是分类器接收的全连接层输出的特征图；通过 softmax 函数计算后输出，得到不同类别的概率值 p_k，输出概率值最大的即为预测类别。

交叉熵损失函数：1D-ICNN 模型中采用损失函数评价输入数据的真实类别与预测类别的一致性。与均方误差相比，交叉熵损失函数更能评估网络模型的质量，因为通过交叉熵运算并不会影响分类函数本身的单调性[14]。交叉熵损失函数的计算为：

$$L=-\frac{1}{m}\sum_{d=1}^{m}\sum_{k=1}^{K}\hat{y}_d^k\log p_d^k \tag{5}$$

式中，m 表示输入数据批量的大小，即为训练样本数据集的数量；K 为类别个数；p_d^k 表示第 d 个数据属于第 k 类的类别预测值；$\hat{y}_d^k$ 表示第 d 个数据属于第 k 类的类别 one-hot 编码真实值。反向传播不断迭代使损失函数的值收敛，求解损失函数对权重和偏置的梯度。

变学习率：学习率越大，模型权重和偏置参数每次更新的程度越大，模型收敛越快；学习率越小，模型权重和偏置参数每次更新的程度越小，模型收敛越慢[15]。为了最小化损失函数，在模型训练初期，保持一段时间较大的学习率可以尽快使网络收敛到最优解附近，可以减小时间开销；在模型训练后期，保持一段时间较小的学习率在最优解附近搜索，可以避免参数在极值两侧跳动，保证了最佳精度。基于衰减学习率变化策略设计自适应动态调整学习率方法为：

$$\alpha(t)=\frac{(\alpha_0-\alpha_e)}{1+\exp(t-T_{med})\times v} \tag{6}$$

式中，α_0 为初始学习率；α_e 为最小学习率；t 为当前迭代次数；T_{med} 为迭代中期次数；v 为预设的正常数。图 2 表示以初始学习率 0.1，迭代次数 100 为例，说明不同学习率衰减策略随迭代次数的变化曲线。

3 老年人自评健康影响因素分析及预测流程

基于实证数据进行研究，首先采用单因素分析中的卡方检验筛选具有统计显著意义的影响因素；然后基于 1D-ICNN 模型进行迭代优化，不断更新网络模型参数，从而对待测老年人自评健康进行预测。老年人自评健康影响因素分析及预测流程见图 3，主要步骤：①输入样本数据，并基于统计学方法进行特征提取，以分析影响老年人自评健康的主要因素；②将特征提取后的老年人自评健康数据划分为训练和测试数据集；③在训练学习过程中，基于本文改进策略，对 1D-ICNN 模型进行前向传播，不断更新网络模型参数。若满足设置的迭代条件，则输出 1D-ICNN 模型结构和参数；否则，基于交叉熵损失函数对 1D-ICNN 模型进行后向传播训练；④在测试学习过程中，将测试数据集输入至训练好的 1D-ICNN 模型中，通过 softmax 层计算预测结果，最终输出老年人自评健康预测结果。

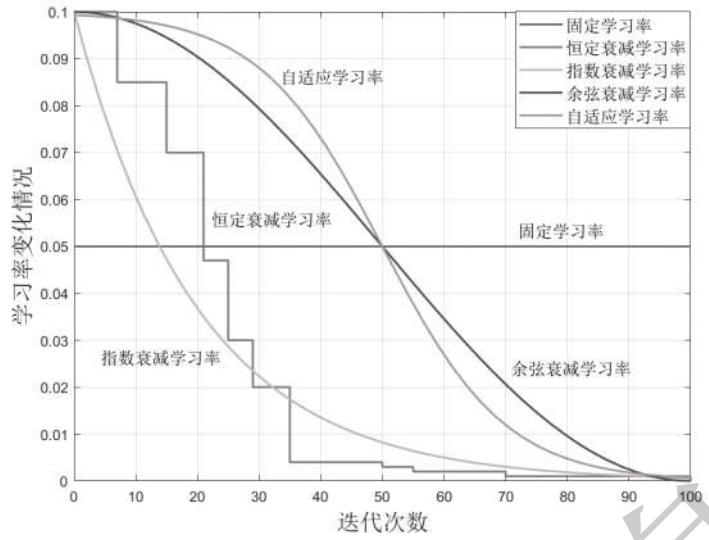

图 2 衰减学习率控制器的变化曲线

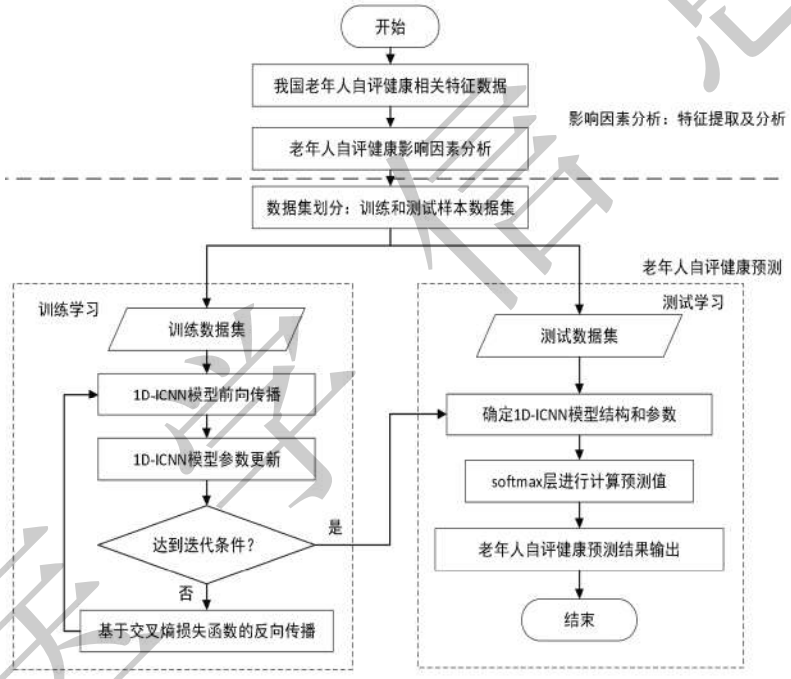

图 3 老年人自评健康影响因素分析及预测流程

4 试验与结果

4.1 老年人自评健康影响因素分析 因变量老年人自评健康以及自变量均属于多分类型变量，因此以卡方检验分析不同因素对于老年人自评健康的显著影响。在假设的显著性水平下（P=0.05），最终筛选出 15 个通过显著性检验的特征变量，见表 2。

在以往的研究中，收入、经济地位、性别、婚姻状况、地区、健康保险、社会参与等因素被认为是影响老年人自评健康的重要因素[16]。就本次结果而言，老年人自评健康状况与职业、睡眠质量、网络信任感知、锻炼频次、就医地点选择、养老居住意愿等方面有关。本研究部分结果与以往研究一致，并且有新发现。

表 2 农村老年人自评健康影响情况

特征变量	χ^2	P	特征变量	χ^2	P
是否少数民族	8.364	0.015	子女情况	24.831	0.000
文化程度	21.939	0.015	是否吸烟	19.041	0.004
政治面貌	10.945	0.027	睡眠质量	62.152	0.000
婚姻状况	25.755	0.004	每周锻炼几次	33.767	0.000
职业	32.603	0.001	在哪里看病	42.117	0.000
年收入	24.661	0.002	网络信任情况	28.788	0.000
年度人情支出	22.460	0.004	您最想在	17.457	0.026
宗教信仰	47.411	0.000	哪里养老		

4.1.1 社会经济地位与老年人自评健康状况的关系 研究发现[17],经济地位对自评健康有直接影响,人均纯收入高的社区自评健康状况好于人均纯收入低的社区。相对贫困与城市和农村老年人的自我评价健康呈负相关[18]。教育程度会对老年人的自评健康产生影响,有研究认为受教育程度较低的人更可能认为自身的健康状况较好,受教育程度较高的人则更可能准确地评价自己的健康状况[19]。本研究表明,相对于其他社会经济的构成因素,职业类别对老年人自评健康程度的影响更为显著。或许是因为职业类别代表了收入和社会地位,以及相对应的各种健康福利。在我国传统的"单位"制度下,职业转变是比较困难的,而在政府机构工作的人员即使在退休后也具有很高的社会影响力,与其他的职业类型相比,他们较高的收入和社会地位对他们的自我健康评估产生了积极影响[20]。

4.1.2 家庭功能与老年人自评健康状况的关系 代际福祉是家庭的基本功能,家庭可以提供对于家庭成员养老的支持。稳定的婚姻关系能够为老年人提供社会支持,从而增强他们的自信心、自我效能感等[21]。大部分学者的研究结果显示有配偶的老年人比没配偶的老年人自评健康状况更好,丧偶、分居或离婚的老年人更容易表现出较差的自评健康[22]。就家庭功能而言,本研究证实了既往研究的观点,老年人对于居住地点的不同偏好对于自评健康有不同影响,倾向于居家养老的老人具有更高的家庭支持水平,因此自评健康状况更好。而倾向于机构集中养老的老人,大部分可能由于家庭支持水平较低而做出的偏好选择,因而更容易产生较差的自评健康状态。

4.1.3 社会参与与老年人自评健康状况的关系 社会参与是"积极老龄化"的重要内容,一般来说,与他人有更多社会互动的老年人,能够得到更多的社会支持,其自评健康状况较好[23]。本研究发现,网络信任感知作为一个社会参与和社会认同的指标,能够一定程度上反应老年人对于社会的亲近感。对于网络社会交往信任水平越高的老年人,自评健康状况越好。而对于老年人来说,锻炼也是社会交往的一种形式,我国老年人经常通过太极拳认识伙伴,扩展社会交往,经常参加锻炼活动的老年人不仅改善了身体功能,还因为锻炼带来的社会支持而有更好的自评健康水平。睡眠质量是一个非常主观的评价因素,睡眠持续时间长不一定会提升自评健康状况,但当老年人认为睡眠质量较差时,通常会带来较低的自评健康状况[24]。

4.2 基于1D-ICNN的老年人自评健康预测结果分析

4.2.1 评估指标 关于老年人自评健康预测问题使用的实证数据集分配规则,选取70%的样本数据进行模型构建和优化,30%的样本数据作为模型预测性能的评估。采用准确率(Accuracy)、精准率/查准率(Precision)、召回率/查全率(Recall)、特异度(Specificity)、AUC(Area Under ROC Curve)指标、运行时间来评估预测性能。

Accuracy:衡量预测正确的结果占所有结果的比例。

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Precision:衡量预测为正的结果有多少实际为正。

$$Precision=\frac{TP}{TP+FP} \quad (8)$$

Recall:衡量实际为正的结果有多少预测为正。

$$Recall=\frac{TP}{TP+FN} \quad (9)$$

Specificity:所有负类数据中被预测正确的比例。

$$Specificity=\frac{TN}{FP+TN} \quad (10)$$

AUC:ROC曲线下的面积即为AUC。

$$AUC=\frac{2}{K(K-1)}\sum_{i=1}^{K}AUC_i \quad (11)$$

式中,TP表示将实际为正类划分为正类的个数;TN表示将实际为负类划分为负类的个数;FN表示将实际为正类划分为负类的个数;FP表示将实际为负类划分为正类的个数;K表示类别个数。

本次采用的实验平台为PyCharm,开发语言为Python3.7,深度学习框架为Keras,该框架可以进行模型的设计、训练、优化以及可视化。训练模型的硬件环境为AMD Ryzen 7 5800H with Radeon Graphics,主频3.20 GHz,内存为16 GB;软件平台为64位Windows10操作系统。1D-ICNN模型训练时,相关参数设置如下:批处理个数为1,epochs为200,激活函数为ReLU,损失函数为交叉熵损失函数,使用自适应时刻估计算法作为优化器,学习率设为自适应衰减变化,初始学习率为0.001,其余参数保持默认值。

4.2.2 预测模型性能分析 为了验证1D-ICNN模型

对老年人自评健康的预测性能，将该算法与传统机器学习算法进行对比，包括：逻辑回归（Logistic Regression, LR）、K 近邻（K Nearest Neighbor, KNN）、支持向量机（Support Vector Machine, SVM）、决策树（Decision Tree, DT）、随机森林（Random Forest, RF）、强分类器（Adaptive Boosting, AdaBoost）、梯度提升决策树（Gradient Boosting Decision Tree, GBDT）、XGBoost（eXtreme Gradient Boosting）、LightGBM（Light Gradient Boosting Machine）、1D-CNN。不同分类器预测性能评估结果见表 3。

表 3 不同分类器预测性能评估结果

算法	Accuracy	Precision	Recall	Specificity	AUC	运行时间(s)
LR	0.561	0.332	0.333	0.673	0.547	0.023
KNN	0.559	0.341	0.345	0.666	0.514	0.004
SVM	0.572	0.359	0.351	0.684	0.601	0.007
DT	0.553	0.328	0.341	0.657	0.500	0.002*
RF	0.560	0.348	0.367	0.657	0.601	0.141
AdaBoost	0.560	0.337	0.351	0.662	0.574	0.094
GBDT	0.554	0.337	0.334	0.666	0.547	3.436
XGBoost	0.574	0.358	0.378	0.670	0.588	0.133
LightGBM	0.575	0.374	0.366	0.683	0.588	0.047
1D-CNN	0.602	0.376	0.435	0.685	0.615	28.924
1D-ICNN	0.794*	0.535*	0.561*	0.692*	0.766*	30.261

注：* 表示最优评估指标值。

由表 3 可以看出，LR 是线性模型，可解释性强，但学习能力有限，需要大量的人工特征工程；KNN 和 SVM 属于涉及到对样本距离度量的模型，如果缺失值处理不当，会导致模型预测效果很差；DT 基础决策树容易产生过拟合的情况，在训练集上有很好的预测精度，在测试集上效果不明显；RF 的随机抽样使得树与树之间没有太多关联性，可能导致拟合效果达到瓶颈；AdaBoost 利用了弱分类器进行级联，考虑了每个分类器的权重；GBDT 训练时间比较长，通常不适用于高维稀疏数据；XGBoost 计算效率高，使用了二阶导，而且有正则化，减少了过拟合；LightGBM 在保证和 XBGoost 精度相当的前提下，提升了速度；1D-CNN 在高维度特征数据预测方面，相较于传统机器学习算法，提高了模型的特征提取能力和学习能力；通过对原始 1D-CNN 模型的优化，采用交叉熵损失函数和变学习率进行网络训练，相较于原始 1D-CNN，1D-ICNN 模型在测试集上的 Accuracy、Precision、Recall、Specificity、AUC 评估指标均为所比较算法中的最优值，明显优于传统机器学习算法，但该算法增加了运行时间。因此，在老年人自评健康预测问题中，该模型特征学习能力较强，提高了预测精度。

5 总结

我国对于农村老年人自评健康的研究存在着一定地域上的局限性。湖南作为较为典型的低城镇化率、高农村人口老龄化率的中部人口净流出省，在这一领域有着很高的研究价值。本文以年龄大于等于 60 岁的老年人作为研究对象，对 2022 年岳阳县老年人养老需求调研数据进行整理，从中提取可能对老年人自评健康产生影响的因素。首先对提取的相关特征因素进行卡方检验，全面探索各个因素对老年人自评健康的影响显著性；然后在此基础上建立老年人自评健康预测模型并进行验证。

本次调研的 369 名农村老年人的自评健康状况不容乐观，多数老年人对自身健康状况评价一般或较差。通过研究发现，文化程度、政治面貌、婚姻状况、职业、年收入等 15 个特征变量与农村老年人自评健康相关；基于筛选的显著影响因素，提出基于交叉熵和变学习率的 1D-ICNN 模型用于预测农村老年人自评健康，通过与传统机器学习算法相比，包括 LR、KNN、SVM、DT、RF、AdaBoost、GBDT、XGBoost、LightGBM、1D-CNN，最终结果表明 1D-ICNN 模型在 Accuracy、Precision、Recall、Specificity、AUC 评估指标上优于所比较的算法。

本文创新之处：①研究视角：从我国农村老年人角

度探究影响自评健康的主要因素，可以更加针对性地解决农村老年人健康问题；②研究数据：通过问卷调查对农村老年人自评健康进行研究，一定程度上解决了缺乏微观层面数据的问题，补充了县域一级具有代表性的实证研究；③研究方法：在卡方分析对老年人自评健康影响因素显著性分析的基础上，进一步采用改进的1D-ICNN模型对老年人自评健康进行预测，突破了传统机器学习分类算法的不足，提供了可应用于识别和预测老年自评健康的深度学习模型。

本文不足及改进：①在对老年人健康状况的分析过程中，没有探究自变量之间交互作用对老年人健康状况的影响，在今后研究中，可以探究多变量之间的交互作用对老年人自评健康的影响，从而更加充实老年人的健康影响因素分析。②本文建立的老年人自评健康预测模型，预测精度还有改进提升空间，可能是由于特征类别数据比较分散，或是仅采用了老年人基本信息、家庭状况、生活方式、网络使用情况、养老需求共五个方面相关特征数据。但在健康中国大背景下，有待于添加更多的相关影响因素来建立预测模型，提高模型的预测精准性。

参考文献：

[1]冉思燕.影响老年旅游者消费水平的因素研究——以重庆市主城区为例[D].重庆:西南大学,2010.

[2]张俊丽,温丹丹,陈素娜,等.横琴65岁以上老年人参加免费健康体检的现状调查[J].医学信息,2023,36(11):81-85,94.

[3]王辉,莫合德斯·斯依提,樊琼玲,等.乌鲁木齐农村老年人养老服务现状分析[J].医学信息,2021,34(6):142-145.

[4]谷琳,乔晓春.我国老年人健康自评影响因素分析[J].人口学刊,2006(6):25-29.

[5]Zadworna-Cieslak M.The measurement of health-related behavior in late adulthood: the health-related behavior questionnaire for seniors[J].Roczniki Psychologiczne,2017,20(3):599-617.

[6]Janszky I,Ljung R,Ahnve S,et al.Alcohol and long-term prognosis after a first acute myocardial infarction: the SHEEP study[J].European Heart Journal,2008,29(1):45-53.

[7]谷景亮.山东省老年慢性病患者用药行为及依从性研究[D].济南:山东大学,2019.

[8]Tomioka K,Kurumatani N,Hosoi H.Association between the frequency and autonomy of social participation and self-rated health [J].Geriatrics and Gerontology International,2017,17(12):2537-2544.

[9]王超,姜茂敏,沈世勇,等.上海市老年人健康素养的城乡差异及影响因素[J].中国卫生事业管理,2023,40(2):148-152.

[10]王可,赵华硕,张虹,等.基于SMOTE算法与机器学习的老年人健康素养预测研究[J].中国校医,2019,33(9):641-643,699.

[11]Cho K,Raiko T,Ilin A.Enhanced gradient for training restricted boltzmann machines[J].Neural Computation,2013,25(3):805-831.

[12]Li Y,Fu Y,Li H,et al.The improved training algorithm of back propagation neural network with self-adaptive learning rate[C]//Proc of Computational Intelligence and Natural Computing.Piscataway,NJ:IEEE Press,2009:73-76.

[13]葛君伟,涂兆昊,方义秋.基于融合CNN和Transformer的分离结构机器翻译模型[J].计算机应用研究,2022,39(2):432-435.

[14]Fei R,Yao Q,Zhu Y,et al.Deep learning structure for cross-domain sentiment classification based on improved cross entropy and weight[J].Scientific Programming,2020,2020:1-20.

[15]Cheng K,Tao F,Zhan Y,et al.Hierarchical attributes learning for pedestrian re-identification via parallel stochastic gradient descent combined with momentum correction and adaptive learning rate[J].Neural Computing and Applications,2020,32(10):5695-5712.

[16]杜本峰,穆跃瑄,盛见.老年人自评健康的预测因素、贡献程度及群体差异[J].中国卫生政策研究,2022,15(4):8-16.

[17]Zadworna M.Pathways to healthy aging-exploring the determinants of self-rated health in older adults[J].Acta Psychologica,2022,228:103651.

[18]Qin W,Xu L,Wu S,et al.Income,relative deprivation and the self-rated health of older people in urban and rural China[J].Front Public Health,2021,9:658649.

[19]Choi A,Cawley J.Health disparities across education:the role of differential reporting error [J].Health Economics,2018,27(3):e1-e29.

[20]He L,Wang K,Wang J,et al.The effect of serving as a danwei leader before retirement on self-rated post-retirement health: empirical evidence from China [J].BMC Public Health,2022,22(1):573.

[21]Knöpfli B,Cullati S,Courvoisier DS,et al.Marital breakup in later adulthood and self-rated health: a cross-sectional survey in Switzerland [J].International Journal of Public Health,2016,61:357-366.

[22]Rana GS,Shukla A,Mustafa A,et al.Association of multi-morbidity, social participation, functional and mental health with the self-rated health of middle-aged and older adults in India: a study based on LASI wave-1[J].BMC Geriatr,2022,22(1):675.

[23]Matud MP,García MC,Fortes D.Relevance of gender and social support in self-rated health and life satisfaction in elderly Spanish people [J].International Journal of Environmental Research and Public Health,2019,16(15):2725.

[24]Coombe AH,Epps F,Lee J,et al.Sleep and self-rated health in an aging workforce [J].Workplace Health and Safety,2019,67(6):302-310.

收稿日期：2023-08-05；修回日期：2023-08-25
编辑/成森